4Т

УДК 004.932.751

*А.С. Тернов*
Інститут кібернетики ім. В.М. Глушкова НАН України, м. Київ, Україна
anton.ternov@gmail.com

# Поскладово-віземний синтез зовнішньої артикуляції для задачі комп'ютерного відтворення української жестової мови

У науковій статті пропонується підхід до синтезу зовнішньої артикуляції для задачі комп'ютерного відтворення української жестової мови на тривимірній моделі голови людини без і з урахуванням особливості фонетичної структури словоформи. Підхід дозволяє відтворювати артикуляцію губ синхронно до анімації жесту з використанням бази морфів візем української мови. Наведена програмна реалізація алгоритму підтвердила працездатність запропонованого підходу.

## Вступ і постановка задачі

Жестова мова зазвичай супроводжується артикуляцією губ, яка, разом з емоційними проявами на обличчі, є допоміжним каналом передачі інформації [1]. Так, для людей з вадами слуху розуміння мовної інформації покращується при можливості бачити обличчя співрозмовника. Дослідження [2] показали, що це справедливо не тільки для обличчя реальної людини, а і для синтезованого обличчя аватару, споглядання імітації артикуляції якого поліпшує рівень сприйняття природної мови в умовах з низьким рівнем співвідношення сигнал-шум. Тому для правильної інтерпретації синтезованої жестової мови необхідно враховувати весь інформаційний комплекс, який супроводжує жест. Окрім цього, однією з проблем при спілкуванні глухих з іншими людьми є вміння сприймати та розуміти усне мовлення, оскільки звичайні люди загалом не знають і не вивчають мову жестів. З цього погляду задача синтезу мімічних та артикуляційних проявів на обличчі є альтернативою мовного спілкування для людей з вадами слуху.

Питання анімації артикуляції тісно пов'язані з задачами розробки аудіовізуальних систем. Але, незважаючи на суттєве просунення у вирішенні задач синтезу візуальної складової мовного процесу, комплексно проблема залишається невирішеною, тому що існуючі підходи мають локальний характер відповідно до задачі та предметної області і є мовозалежними. Так, розробкою методів моделювання міміки займаються у багатьох країнах світу, зокрема в Америці [3], Швеції [4], Англії [5], Німеччині [6], [7] та інших [8], [9], [10]. Більшість робіт спрямовані на задачі створення аудіовізуальних систем для розробки локалізацій інтелектуальних комп'ютерних інтерфейсів та інтерфейсів для мобільних телефонів з синтезом зорової складової мовного процесу людини (деякі з них безпосередньо присвячені питанням інтеграції людей з вадами слуху у суспільство [5], [8]). Особливості мовленнєвого процесу української мови, зумовлені відмінностями як у фонетиці, так і морфології, вимагають перегляду та адаптації існуючих підходів до задачі відтворення зовнішнього артикуляційного процесу при синтезі української жестової мови. З цього погляду дослідження І.К. Білодіда [11] з тематики артикуляційних особливостей при промовлянні у сучасній українській літературі та мові можуть бути покладені в основу підходу до синтезу міміки та артикуляції в українській жестовій мові. А в силу подібності за фонетикою російської мови до української, цінність мають роботи А.Л. Воскресенського [12], які направлені на створення цифрового жестівника

російської жестової мови. Слід також відзначити результати досліджень [13], які вилились у рекомендації з анімації мімічних проявів та артикуляції при відтворенні розмовної англійської та російської мови з використанням 3D-моделі голови людини.

Крім проблем, пов'язаних з локалізацією мови, мають місце і такі чинники:

– при відтворенні жестової мови існують деякі особливості зовнішньої артикуляції (наприклад, артикуляція, яка супроводжує жестові одиниці, чи частково, чи зовсім не пов'язана з розмовною мовою) [14];

– внаслідок лише логічної відповідності між жестом та фонетичною структурою слова, яке має анімуватися артикуляцією, часові проміжки для жестової анімації і анімації процесу артикуляції чуючої людини можуть відрізнятьcя досить суттєво [15].

Саме тому, для природності відтворення артикуляції при синтезі жестової мови, на відміну від задач аудіовізуального синтезу, необхідно провести синхронізацію артикуляції з анімацією жесту. На практиці сурдоперекладачі природно сповільнюють процес вимови на обличчі, коли можуть виникнути запізнення за часом процесу показу жесту (час для показу жесту більший, ніж час для артикуляції слова в нормальному темпі).

Проведений аналіз існуючих робіт з даної тематики визначив напрямок досліджень і постановку задачі.

**Постановка задачі.** Необхідно синтезувати природну артикуляцію для процесу візуалізації української жестової мови на тривимірній моделі людини з синхронізацією часових проміжків анімації артикуляції та анімації жесту.

– підхід до синтезу має узгоджуватись з загальною концепцією технології невербального спілкування людей з вадами слуху;

– підхід має враховувати особливості артикуляції при відтворенні української мови.

Вважається, що кожен жест має своє слово-образ (слово чи набір слів української мови у нормальній словоформі).

## Синтез жестової мови

Задача синтезу візуальної складової артикуляційного і мімічного процесу є складовою задачі синтезу жестової мови і буде залежати від підходу до реалізації останньої.

Можна виділити два принципово різних підходи до синтезу жестової мови:

1. Статичний. Прикладами можуть слугувати жестівники та тлумачі жестової мови з набором жестових одиниць та сталих виразів. Із реченням чи словом розмовної мови зіставляється його аналог у жестовій мові. Формат даних: паперовий, відеодиски, програмні продукти (прикладні програми, інтернет-плагіни та сервіси), які використовують відеофайли та анімовані зображення [12]. Внаслідок специфіки формату даних та складності реалізації плавності переходів такий підхід не дозволяє генерувати чи синтезувати жестові речення та нові жестові одиниці окрім тих, що містяться в базі.
2. Динамічний. Формат даних: відцефровані чи змодельовані рухи та міміка людини – носія жестової мови. При такому підході є можливість використовувати жестові одиниці для продукування жестової мови [5], [16]. Може використовуватись в системах синхронного чи асинхронного сурдоперекладу текстової, мовної чи відеоінформації, для створення інтелектуальних інтерфейсів для людей з вадами слуху.

У зв'язку з розвитком цифрових технологій та потужності (продуктивності) комп'ютерної техніки особливу цікавість становить другий підхід, який дозволяє створювати інтелектуальні інтерфейси та гіпермедійні технології і засоби для розробки навчальних систем і систем штучного інтелекту.

Основні проблеми і задачі динамічного підходу:

– для отримання можливості динамічного відтворення жестової мови необхідно мати її формальний опис чи представлення, як, наприклад, для BSL (British Sign Language) та DGS (German Sign Language) у форматі Гамбурзької системи нотації

(HamNoSys language) [10]. Формальний опис використовується для розробки алгоритмів та підходів до синтезу речень жестової мови з жестових одиниць. Необхідною для розв'язання є задача моделювання людиноподібних переходів при конкатенації елементів візуального синтезу [4], [9].

– моделювання людиноподібного аватару. Використання аватару з високою деталізацією для моделювання рухів людини з анімацією міміки та артикуляції і врахуванням природних для людини рухів голови під час розмови жестовою мовою [16], [17].

Тому в роботі [12] відмічається складність створення аватару для продукування жестової мови.

В Україні поки що не створено повного формального опису для української жестової мови, який задовольняв би наведеним вимогам. Розробками в даному напрямку займаються у лабораторії жестової мови [18], але дослідження мають методологічний характер і направлені на розробку теоретико-методичних засад використання жестової мови у навчанні нечуючих та створенні навчальних курсів та методичних посібників для навчального процесу з вивчення української жестової мови. Тому для розв'язання задачі моделювання людиноподібного аватару в роботі [16] авторами запропонована інформаційна технологія невербального спілкування для людей з вадами слуху, яка містить у собі функціональність з синтезу: рухів жестової мови, дактильної абетки, артикуляційної й емоційної міміки на тривимірній моделі людини. Елементами для синтезу жестової мови виступає множина відцифрованих жестів та відповідних їм нормальних форм слів розмовної мови. Формальна модель жестової мови має спрощену структуру, але враховує основні вимоги щодо подання жестів, емоцій та артикуляції. Приклад речення жестовою і розмовною українською мовами:

«[ФОТО] [!INTORREGATIVE!ХТО]» = «Хто на фото?»

Для реалізації модуля синтезу правильної міміки й артикуляції на тривимірній моделі людини необхідно було вирішити ряд проблем, винесених в постановку задачі даної статті.

# Синтез та синхронізація артикуляції

У задачах анімації мовленнєвого процесу на моделі голови людини існує декілька різних підходів до синтезу процесу артикуляції. Класифікація вживаних методів базується на відмінностях у типі даних, які доступні для синхронізації:

– текстові дані: на вхід системи подається текст, який транскрибується в послідовність фонем. Ця інформація використовується для генерації синтезованої мови та синхронізації артикуляції;

– мовні (голосові) дані: на вхід подається звуковий запис мови. Аналізується аудіофайл для отримання послідовності фонем та часових тривалостей [19];

– комбінований текстово-мовний підхід: текст і його фонетичне подання використовується для пошуку часових границь мовних сегментів у звуковому сигналі для того, щоб отримати інформацію о тривалостях елементів анімації.

Для задачі синхронізації артикуляції з жестом у роботі пропонується модифікація комбінованого підходу, яка використовує текстові дані, а також додаткову інформацію про часову тривалість звучання фонем та їх часові границі для конкретного слова. Цю додаткову інформацію було отримано з даних, занесених в принципи роботи звукового синтезатора, реалізованого у технології [6], який використовує результати досліджень І.К. Білодіда [11], та структури подання синтезатором слова. Інформація, яка використовується для синхронізації, отримується на етапі обробки текстового подання нормальної словоформи відповідно до жесту. Схема процесу обробки наведена на рис. 1.

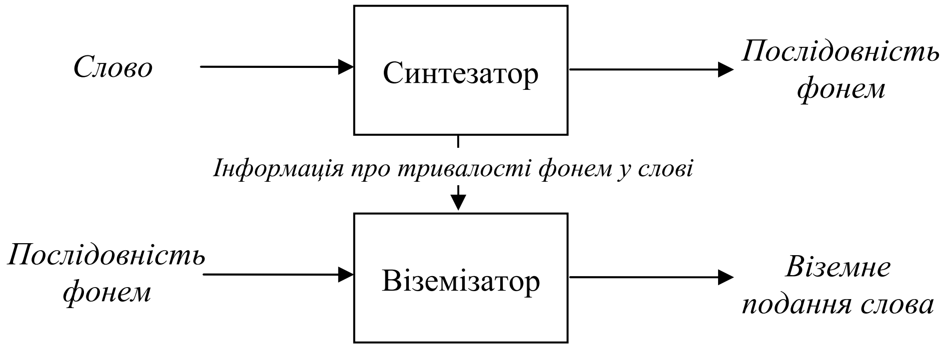



Рисунок 1 – Вхідні дані для різних методів

Приклад роботи віземізатора для вхідної нормальної словоформи «багато» наведено на рис. 2.


**Віземне подання слова:** БАГАТО -> < | ПБМ(7)А(7) | КГХГк(4)А(7) | Т'Д'Н'(5)О(9)>

Рисунок 2 – Приклад роботи віземізатора для слова «багато»: де «ПБМ», «А» та інші – назви візем, «|» – границі сегментів, «(7)» – кількість кадрів анімації без синхронізації для кожної віземи

Елементами блоку «віземізатора» є віземи. У роботах [11], [20] робиться висновок, що візуальний алфавіт мови є істотно неповним. У ньому немає однозначної відповідності між вимовленою фонемою і її візуальним відображенням, що знижує можливості зорового сприйняття мови. Фонеми, які виглядають подібними одна до одної, під час артикуляції можна віднести до однієї групи, яка називається віземою. Тобто елементами візуального алфавіту виступають віземи (віземи для української мови [20]). У проведених дослідженнях використовувалась бібліотека візем для української мови, принципи побудови якої було описано в роботі [15].

Загальна схема процесу синхронізації анімації артикуляції з жестом зображена на рис. 3.

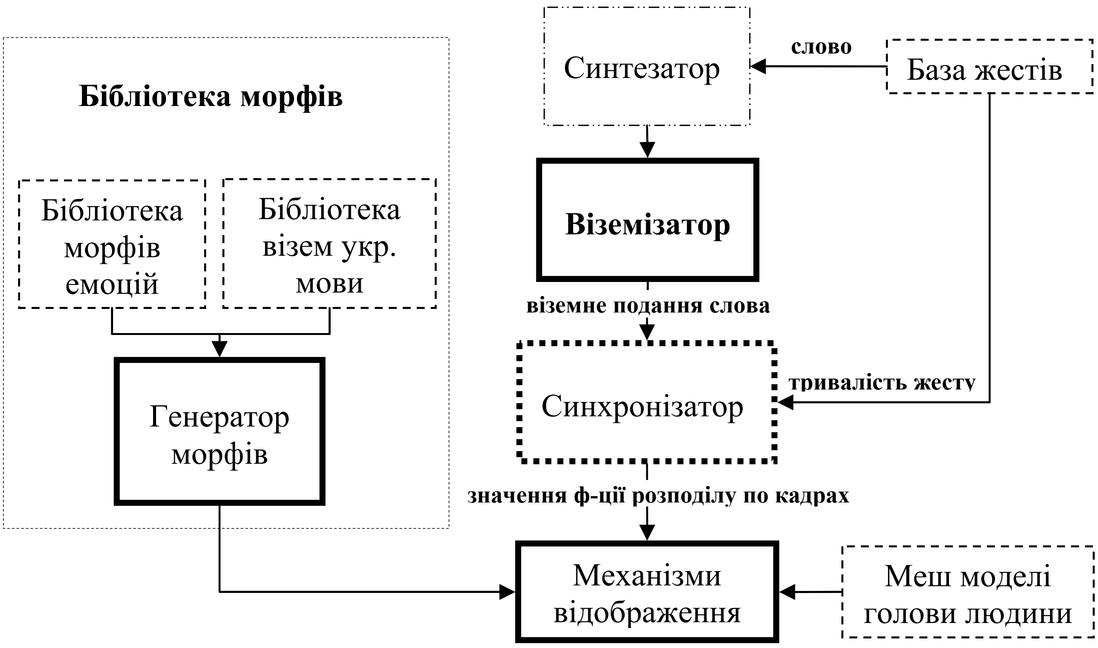



Рисунок 3 – Схема синтезу анімації артикуляції і міміки

Для природності анімації артикуляції пропонується робити синхронізацію шляхом деформування часових границь сегментів, що відповідають голосним фонемам. З деяких міркувань особливості артикуляції щілинних звуків [11], до класу фонем, що можуть змінювати часовий проміжок, також можна включити фонеми, які відповідають віземам «ВФ» та «ШЖЧДж», якщо вони стоять на початку слова і наступним іде приголосний звук.

У дослідженні розглядалось два методи синхронізації з урахуванням і неврахуванням віземної структури слова. Відмінності будуть відображатися на часових діаграмах.

**Повіземний синтез і синхронізація.** Для синхронізації з анімацією жесту змінюються часові проміжки, які безпосередньо відповідають голосним фонемам. Такий підхід мав за мету отримати потрібний результат без додаткових перетворень і врахування фонемної структури слова, бо голосні фонеми мають яскраво виражений візуальний портрет (візему) [11], [13] і зміни в часі звучання не мали би зменшити рівень сприйняття візуальної інформації (в даному випадку – артикуляції).

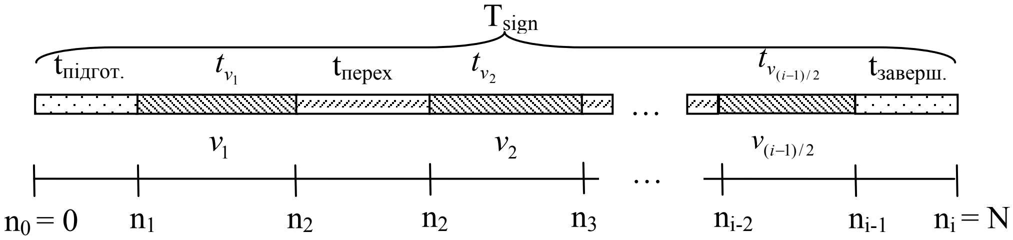



Рисунок 4 – Часова діаграма для повіземного синтезу зі синхронізацією

Нехай $v_l$ – візема, яка стоїть на $l$-у місці у слові-образі (рис. 4), $V^v$ – множина голосних вісем (віземи голосних фонем), $V^c$ – множина приголосних віземи, $T_{sign}$ – час показу жесту, $t_i^c$ – тривалість приголосного звуку $i$, $t_j^v$ – тривалість голосного звуку $j$ в слові-образі жесту. Усі мовні конструкції промовлялись у нормальному для людини темпі. Синтезоване слово зображено послідовністю таких тривалостей, наприклад, $(t_1^c, t_1^v, t_2^c, t_3^c, t_2^v)$. Поклавши:

$$T_{sign}^v = \sum_i t_i^v, \quad T_{sign}^c = \sum_j t_j^c, \quad T_{trans} = t_{підг} + t_{заверш} + ((i-1)/2-1)\cdot t_{перех}, \tag{1}$$

$$T_{dif} = T_{sign} - \sum_j t_j^c - T_{trans} = T_{sign} - T_{sign}^c - T_{trans}, \tag{2}$$

де $t_{підг}, t_{заверш}$ – час, необхідний на підготовку м'язів роту для анімації першої віземи, та повернення у стан спокою відповідно, $t_{перех}$ – час переходу між віземами , а $T_{trans}$ – відповідає за загальний час усіх перехідних станів. Час переходу між віземами вважався однаковим для усіх переходів.

Тоді послідовність тривалості візем для синтезу зовнішньої артикуляції матиме такий вигляд: $(t_1^c, p\cdot t_1^v, t_2^c, t_3^c, p\cdot t_2^v)$, де $p = T_{dif} \big/ T_{sign}^v$. Визначаючи частоту появи кадрів $\omega$, можна навести характеристики тривалості в кадрах $k = [t\cdot\omega]$.

У дослідженнях [15] синхронність до анімації жесту завжди мала місце, але при такому підході добре (природно) синтезувались лише слова-образи довжиною 4, 6 фонем з CV-складами («ма-ти», «ба-га-то» і таке інше). Нереалістично відображались складні слова-образи («покласти», «скільки»). Тому було вирішено враховувати фонетичну структуру слова-образу.

**Поскладовий синтез і синхронізація.** Принцип синхронізації, який покладено в основу цього підходу, має за своє підґрунтя вдале використання його аналога в задачах синхронізації анімації з мовним сигналом [4]. Так, кількість фреймів для анімації жесту розподіляється між «складами», а не віземами слова-образу. У проведених дослідженнях розглядався рівномірний розподіл часу. Під складом $s$ розуміється послідовність візем (фонем): $s_k=(v_{l_{k_1}},..,v_{l_{\rho_k}})$, $l_{k_1},..,l_{\rho_k}$ – послідовні індекси візем слова подання $l_{k_{i+1}}-l_{k_i}=1$. Причому для $\forall k^1,k^2$, для складів $s_{k^1},s_{k^2}$ виконується наступне: $l_{k^1_1}\notin[l_{k^2_1},l_{\rho_{k^2}}]$ і $\sum_k \rho_k=(i-1)/2$. Тобто склади не порожні і не перетинаються. $S$ – множина складів слова-образу.

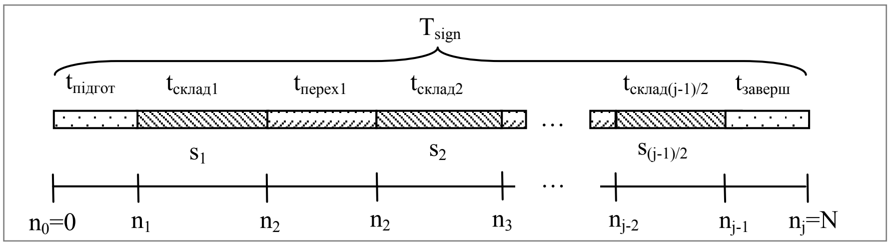



Рисунок 5 – Часова діаграма для поскладового синтезу з синхронізацією

Особливості перетворення слова синтезатором у фонемне подання (рис. 1) можуть призвести до випадків появи «складів» без голосної фонеми. Нехай «склад» називається повним, коли в ньому присутня голосна фонема. Множина повних складів позначатиметься $S^{complete}=\{s_m \mid \exists q: v_{l_{m_q}}\in V^v\}$, $S^{incomplete}=S-S^{complete}$. Нехай:

$$t_{s_m}=\sum_{k=1}^{\rho_m} t_{v_{l_{m_k}}},\ T_{complete}=\sum_{s_m\in S^{complete}} t_{s_m},\ T_{incomplete}=\sum_{s_m\in S^{incomplete}} t_{s_m}, \tag{3}$$

$$t_{перех_k}=f(t_{склад_{k-1}},t_{склад_k}),\ T_{trans}=t_{підг}+t_{заверш}+\sum_{k}^{(i-1)/2-1} t_{перех_k}, \tag{4}$$

$$T_{dif}=T_{sign}-T_{incomplete}-T_{trans}, \tag{5}$$

де $t_{перех_k}$ залежить від попереднього і наступного складу, $T_{incomplete}$ – час, який би займала анімація неповних складів. Усі базові часові характеристики – це дані, отримані від звукового синтезатора без синхронізації до жесту.

Для врахування синхронізації деформуватися будуть лише повні склади за формулою:

$$\bar{t}_{склад_m}=p\cdot t_{склад_m},\ p=\begin{cases}1, & s_{склад_m}\in S^{incomplete}\\ T_{dif}/T_{complete}, & інше\end{cases}. \tag{6}$$

При такому підході можна застосовувати алгоритми до анімації візуально-голосової мовної інформації (наприклад, [13]) в межах складу, не порушуючи синхронність.

Загальним недоліком обох підходів до синхронізації артикуляції губ з жестом є те, що вони не враховують елементи синтезу при невикористанні чи відсутності відповідного слова-образу. Але це буде несуттєвим, якщо враховувати обмеження задачі перетворення розмовної мови в жестову з використанням лише бази жестових елементів зі словами-образами, як у навчальних посібниках для жестової мови [18]. У дослідженнях множина жестів була також обмежена жестами, які мають слово-образ, що зумовлено відсутністю наповнення бази морфів унікальними для відтворення жестової мови морфами емоційних проявів невербальної природи для передачі додаткової інформації про жест. Для включення їх до розгляду та врахування зазначеного недоліку необхідно провести додаткові дослідження.

# Результати експериментальних досліджень

Для тестування запропонованого підходу було створене відповідне програмне забезпечення мовою C# (рис. 6), яке дозволяє дослідити вплив зміни параметрів моделі на зрозумілість та розбірливість синтезованої зовнішньої артикуляційної складової мовленнєвого процесу при відтворенні жестової мови.

Метою дослідження не була розробка нового 3D-аватару з власною системою м’язів та морфів, тому для тестових прикладів була використана стандартна модель Симон з Poser 7 [16], яка дозволяє реалістично моделювати більшість емоційних проявів на обличчі людини та має функціональність для створення власних морфів з урахуванням рухів основних зовнішніх артикуляторів (щелеп, язика) і яка, на відміну від інших програмних засобів (Maya, 3D Studio Max), була розроблена безпосередньо для комп’ютерного моделювання людини. Усі базові віземи та емоційні стани були отримані за допомогою функціональності Face Morph. Додаткові морфи емоцій було побудовано з використанням тестової версії програмного продукту FaceGen Modeller [3].

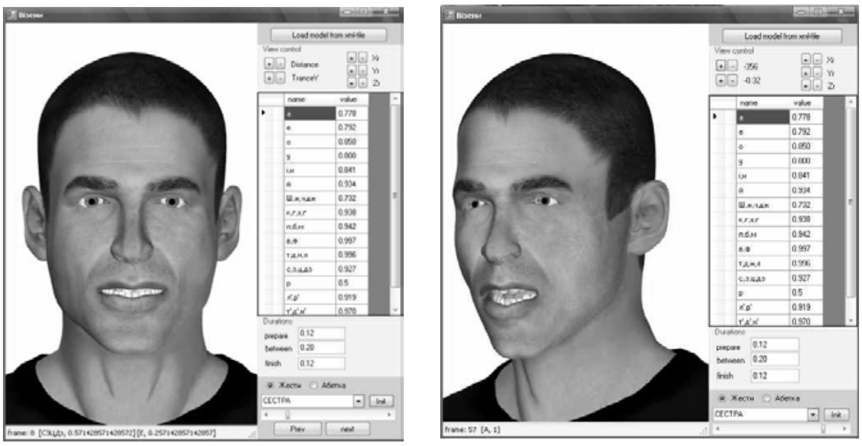
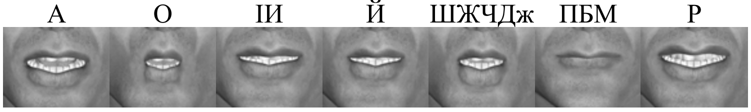


Рисунок 6 – Головне вікно програмної реалізації

Для кодування та передачі інформації про мімічний стан моделі голови було розроблено власний формат даних, який використовується в розробках [16], до складу якого входить опис мешу 3D-моделі з відповідними морфами візем і емоцій, що винесені в окрему бібліотеку для їх незалежного редагування окремо від самої моделі людини.

Таблиця 1 – Морфи візем, створені за допомогою Face Morph

| А | О | ІИ | Й | ШЖЧДж | ПБМ | Р |
|---|---|---|---|---|---|---|



| Е | У | КГХГк | Т'Д'Н' | ФВ | ТДНЛ | СЗЦДз |
|---|---|---|---|---|---|---|



У дослідженнях параметрами системи виступали тривалості переходів між віземами та складами візем, різні варіанти початку та кінця артикуляційного мовлення, різні підходи до синхронізації зовнішньої артикуляції з жестом (тривалості показу візем), але для спрощення $t_{перех_k} = const$.

Для кожного слова-образу, через обмеження на кількість кадрів анімації, наведені параметри варіювалися в залежності від часових характеристик елементів його структури з метою максимізації кількості кадрів для плавної анімації переходу між складами візем ($[t_{пер} \cdot \omega]$) без втрати правильності сприйняття. Для виконання вимоги

синхронізації та невід’ємності часу анімації на формулу 5 накладається обмеження:

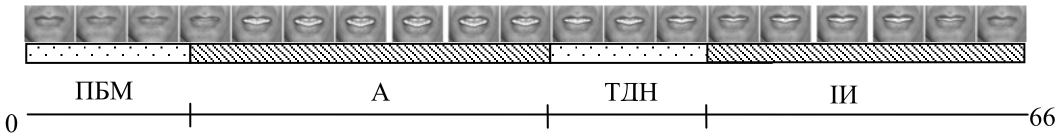
$$T_{dif} > 0\,. \qquad (7)$$

У дослідженнях $\omega = 33$, $t_{підгот}$ і $t_{заверш}$ обирались таким чином, щоб анімація з підготовки та завершення вимови була природною. При виконанні обмеження (7) єдиним додатковим обмеженням параметрів з підготовки і завершення анімації артикуляції буде лише таке: щоб кількість кадрів для їх анімації була не менше трьох [13]. Це робиться для того, щоб досягти мінімальної плавності переходу від і до стану спокою. При великих значеннях цих параметрів аватар буде трохи «втомлено» починати «розмову». Для деяких слів-образів величини параметрів моделі (3 – 6) наведені у табл. 2.

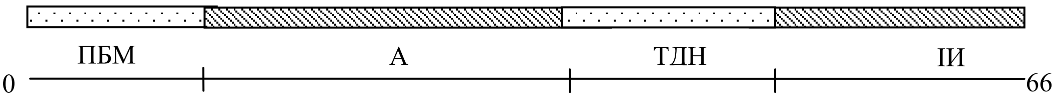
Таблиця 2 – Часові параметри тривалості у мілісекундах

| Час \ слово-образ | «ДАКТИЛЮВАТИ» | «ДАТИ» | «БАТЬКО» |
|---|---|---|---|
| Час початку анімації | 0,091 | 0,15 | 0,12 |
| Час анімації переходу | 0,105 | 0,18 | 0,14 |
| Час кінця анімації | 0,091 | 0,15 | 0,12 |

На тестовій вибірці з 50 жестів поскладова синхронізація анімації слова-образу з жестом виявилась кращою у порівнянні з повіземною синхронізацією без врахування фонетичної структури, бо вдалося врахувати пропорційність часу анімації складу в розмовній мові (точки початку анімації складів розташовуються відповідно до внутрішньої будови складу) та можливість робити додаткові модифікації у способах анімації переходів у складі без зміни положень точок початку анімації складів.



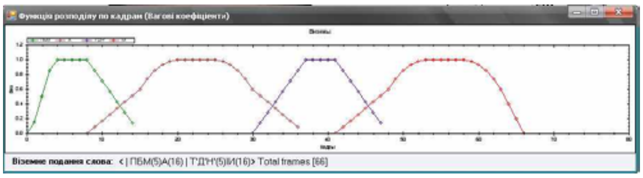

а) з врахуванням складової структури слова



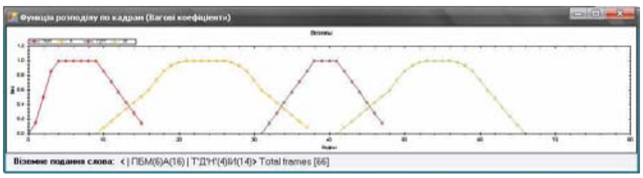

б) без врахування

Рисунок 7 – Приклад послідовності артикуляції слова-образу «мати»

Вагові коефіцієнти показу віз ем обчислюються з використанням функції розподілу за кадрами [15]. Вагові коефіцієнти функції розподілу за кадрами подано графіками на рис. 8



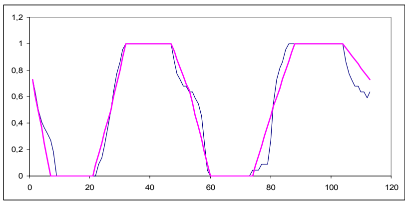
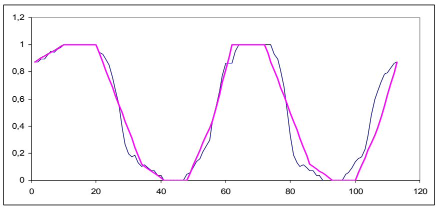
а)



б)

Рисунок 8 – а) з врахуванням складової структури слова, б) без врахування

Для аналізу якості результату синтезу використовувалась суб'єктивна модель оцінки сприйняття, що має за основу оцінку (висновок), отриманий від експерта в галузі сурдоперекладу, та оцінку схожості графіків відхилення чотирьох точок, які відповідають за верхню $T_{top}$, нижню $T_{bot}$ точки роту і кути роту $T_{left}$ і $T_{right}$. Для побудови графіків відносної зміни точок спостереження аналізувались послідовності віднормованих і відорієнтованих зображень областей губ сурдоперекладача і моделі голови людини при відтворенні жестової мови (рис. 9).

а) б)

Рисунок 9 – Для жесту «мама»: а) нормовані відхилення для верхньої точки роту; б) для нижньої точки роту

# Висновки

Наведена програмна реалізація алгоритму поскладово-віземного синтезу підтвердила працездатність запропонованого підходу до синхронізації анімації артикуляції і анімації жесту. Запропонований підхід дозволяє проаналізувати вплив зазначених параметрів часової діаграми анімації слова-образу відповідно до створеної бази візем і емоцій з метою покращення останньої за рахунок розширення чи заміни її складових елементів на ті, які поліпшать візуальне сприйняття зовнішньої артикуляції в деяких переходах між віземами чи «мовними» конструкціями (складами).

Подальші дослідження будуть спрямовані на:

– на врахування різних типів коартикуляції в складах візем української мови для покращення природності візуального сприйняття;

– на використання нерівномірного розподілу часу між складами в залежності від його структури і його фонетичних властивостей (наголошеності чи ненаголошеності і таке інше);

– на розробку алгоритмів, які будуть враховувати особливості впливу артикуляції наступного звуку іншого складу на попередній.

Також планується включити до розгляду мімічні прояви та артикуляцію, яка є унікальною для жестової мови.

Задачі синтезу тісно пов'язані з класом задач з розпізнавання зорових образів і створення навчальних програм для чуючих людей з метою покращення артикуляції губ для їх більш точного розпізнавання. Тому дана робота також мала за мету отримати нові знання і додаткову інформацію про природу та механізми артикуляції губ людини при відтворенні жестової мови для виявлення прихованих параметрів. Отримані параметри будуть використані для проведення подальших досліджень з розпізнавання зовнішньої артикуляції мовного процесу [20] з метою створення системи розпізнавання та аналізу мімічних проявів на обличчі людини – носія жестової мови.

# Література

1. Чистович Л.А. Речь. Артикуляция и восприятие / Л.А. Чистович, В.А. Кожевников. – М. ; Л. : Наука, 1965. – 242 с.
2. Beskow J. The Teleface project – disability, feasibility and intelligibility [Електронний ресурс] / [J. Beskow, M. Dahlquist, B.Granström, M. Lundeberg, T. Öhman]. – Режим доступу : http://www.speech.kth.se/~beskow/papers/fon97teleface.pdf.
3. Parke F. Computer Facial Animation / Frederic I. Parke, Keith Waters. – A.K. Peters Ltd. Wellesley, 2008. – 454 p.

4. Beskow J. Trainable Articulatory Control Models for Visual Speech Synthesis / J. Beskow // Journal of Speech Technology. – 2004. – № 7(4) – P. 335-349.
5. Glauert J. Virtual Human Signing as Expressive Animation / [J. Glauert, R. Kennaway, R. Elliott, B.-J. Theobald] // In Symposium on Language, Speech and Gesture for Expressive Characters. – University of Leeds, 2004. – P. 98-106.
6. Damien P. Phoneme-Viseme Mapping for German Video-Realistic Audio-Visual-Speech-Synthesis [Електронний ресурс] / P. Damien, N. Wakim, M. Egea – Режим доступу : http://www.sk.uni-bonn.de/forschung/ikp-arbeitsberichte-neue-folge/ikpab-nf11.pdf
7. Tümmler J. Avatare in Echtzeitsimulationen / Jörn Tümmler // Master thesis, Kassel University Press. – Germany, December. – 2007. – 192 p.
8. Verlinden M. Multimedia with Animated Sign Language for Deaf Learners / M. Verlinden , I. Zwitserlood, Frowein H. // In P. Kommers & G. Richards (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications. – Montreal, Canada, 27 June 2005. – P. 4759-4764.
9. .Belkowska J. Audiovisual Synthesis of Polish Using Two- and Three-Dimensional Animation / J. Belkowska, A. Glowienko, K. Marasek // Computer vision and graphics: International conference, ICCVG 2004, Warsaw. – 2004. – Vol. 32. – P. 1082-1087.
10. Kennaway R. Experience with and requirements for a gesture description language for synthetic animation / R. Kennaway // Lecture Notes in Computer Science. – 2004. – Vol. 2915. – P. 300-311.
11. Білодід І.К. Сучасна українська літературна мова / Білодід І.К. – К. : Ін-т мовознавства ім. О.О. Потебні ; Наукова думка, 1969. – 435 с.
12. Воскресенский А.Л. От звучащей речи к жестовой / А.Л. Воскресенский, Г.К. Халагин // Речевые технологии. – 2009. – № 1. – С. 99-106.
13. Флемінг Б. Методы анимации лица. Мимика и артикуляция / Б. Флемінг, Д. Доббс ; [пер. с англ.]. – М. : ДМК Пресс, 2002. – 336 с.
14. Жестова мова й сучасність: збірник наукових праць, випуск 3 / ред. кол., головний ред. В.В. Засенко. – К. : Педагогічна думка, 2008. – 192 с.
15. Крак Ю.В. Синтез зовнішнього артикуляційного процесу на обличчі людини для моделювання жестової мови / Ю.В. Крак, А.С. Тернов // Журнал обчислювальної та прикладної математики. – 2009. – № 3(99). – С. 48-56.
16. Кривонос Ю.Г. Інформаційна технологія для моделювання української мови жестів / Ю.Г. Кривонос, Ю.В. Крак О.В. Бармак [та ін.] // Искусственный интеллект. – 2009. – № 3. – С. 186-197.
17. Gaglio S. A Java3D Talking Head for a Chatbot / [Salvatore Gaglio, Giovanni Pilato, Roberto Pirrone, Orazio Gambino, Agnese Augello, Alessandro Caronia] // International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2008). – Barcelona, Spain, 4 – 7 March, 2008. – P. 709-714
18. Адамюк Н.Б. Програма-комплекс «Українська жестова мова» [Електронний ресурс] / Адамюк Н.Б. – Режим доступу : http://www.mon.gov.ua/main.php?query=education/average/programs_ gluh. – Назва з екрана.
19. Kohnert C.J. Lip Synchronization Using Linear Predicitve Analysis / C.J. Kohnert, Sudhanshu Kumar Semwal // Proceedings of the IASTED International Conference on SignalProcessing, Pattern Recognition, and Applications, SPPRA2006. – Innsbruck, Austria, 15 – 17 February, 2006. – P. 210-215.
20. Крак Ю.В. Информационная технология для автоматического чтения по губам украинской речи / Ю.В. Крак, О.В. Бармак, А.С. Тернов // Комп'ютерна математика. – 2009. – № 1. – С. 86-95.

***А.С. Тернов***

**Слогово-виземный синтез внешней артикуляции для задачи компьютерного воспроизведения украинской жестовой речи**

В научной статье предлагается подход к синтезу внешней артикуляции для задачи компьютерного моделирования украинской жестовой речи на трехмерной модели головы человека с учетом особенностей фонетического строения слова-образа. Подход позволяет отображать синтезируемую анимацию артикуляции губ синхронно с анимацией жеста. Для синтеза используется база морфов визем украинского языка. Программная реализация алгоритма предложенного подхода показала его работоспособность.

***A.S. Ternov***

**Syllable-Viseme Synthesis of External Articulation for the Problem of Computer Reproduction of Ukrainian Sign Language**

An approach to the synthesis of external articulation for the problem of computer reproduction of the Ukrainian sign language in a three-dimensional model of a human head with and without taking into account features of phonetic structure of word forms is proposed in this paper. The approach allows one to reproduce the articulation of lips synchronously with gesture animation using a database of morphs of the Ukrainian language visemes. The implementation of the software of an algorithm described proves functionality of the proposed approach.

*Стаття надійшла до редакції 19.07.2010.*